Genomics and proteomics: Applications in autoimmune diseases

Wolfgang Hueber1,2,3
William H Robinson1,2
1VA Palo Alto Health Care System, Palo Alto, CA, USA; 2Division of Immunology and Rheumatology, Stanford University School of Medicine, Stanford, CA, USA; 3Novartis Institutes of Biomedical Research, Novartis, Basle, Switzerland

Abstract: Tremendous progress has been made over the past decade in the development and refinement of genomic and proteomic technologies for the identification of novel drug targets and molecular signatures associated with clinically important disease states, disease subsets, or differential responses to therapies. The rapid progress in high-throughput technologies has been preceded and paralleled by the elucidation of cytokine networks, followed by the stepwise clinical development of pathway-specific biological therapies that revolutionized the treatment of autoimmune diseases. Together, these advances provide opportunities for a long-anticipated personalized medicine approach to the treatment of autoimmune disease. The ever-increasing numbers of novel, innovative therapies will need to be harnessed wisely to achieve optimal long-term outcomes in as many patients as possible while complying with the demands of health authorities and health care providers for evidence-based, economically sound prescription of these expensive drugs. Genomic and proteomic profiling of patients with autoimmune diseases holds great promise in two major clinical areas: (1) rapid identification of new targets for the development of innovative therapies and (2) identification of patients who will experience optimal benefit and minimal risk from a specific (targeted) therapy. In this review, we attempt to capture important recent developments in the application of genomic and proteomic technologies to translational research by discussing informative examples covering a diversity of autoimmune diseases.

Keywords: proteomics, genomics, autoimmune diseases, antigen microarrays, 2-Dih, rheumatoid arthritis, Crohn’s disease, SLE, multiple sclerosis, GWAS

Introduction

About 3%–5% of all human diseases are classed as autoimmune diseases, and these conditions are typically heterogeneous in their clinical presentations and disease courses and outcomes. Great unmet medical need exists for (i) improved diagnosis and outcome prediction of, and (ii) innovative therapies for, complex autoimmune diseases such as multiple sclerosis (MS), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), systemic sclerosis, and Crohn’s disease (CD). An individualized pharmacotherapy that results in sustained remission rather than mere symptom relief is considered the ultimate goal for the long-term management of patients with autoimmune disease.

Great progress has been made in the field of genomic and proteomic technologies for the identification of novel drug targets and molecular signatures associated with clinically important disease states, disease subsets, or differential responses to therapies. In parallel, the elucidation of cytokine networks and the development of pathway-specific biological therapies revolutionized the treatment of autoimmune diseases.
In this review, we examine recent genomic and proteomic studies of clinical relevance, and provide examples of the following: (a) identification of new pathways and novel molecular targets relevant to drug discovery and development; (b) molecular classification of heterogeneous autoimmune diseases; (c) identification of susceptibility markers and “at-risk” individuals; and (d) prediction of disease outcome and response to therapy. Several excellent reviews that provide technology and methods overviews or detailed discussion of genome-wide association studies (GWAS) in autoimmune diseases were recently published, and studies covering these topics will only be highlighted where appropriate for the scope of this review.

While many of the breakthroughs in the development of high-throughput technologies for the genome-wide genotyping and proteomic phenotyping of patients occurred a decade or more ago, larger-scale autoimmune disease studies that apply these technologies to screen relevant patient populations and validate the findings in independent patient cohorts are just starting to be performed.

Discoveries of molecular signatures associated with susceptibility, severity, and long-term outcomes of disease and with distinct treatment requirements have been reported for major autoimmune diseases over the past few years, and some of these results are discussed below. Over the same period, a limited number of groundbreaking exploratory investigations have used genomic and proteomic techniques to identify fundamentally novel targets for the pharmacotherapy of autoimmune diseases.

**Discovery of novel therapeutic targets by genomic and proteomic analysis of human tissue**

Examples of innovative and successful approaches to drug target discovery are provided by studies on MS. Novel targets for therapeutic intervention in MS were recently discovered by transcriptional analysis of active MS lesions. To select candidate targets for subsequent confirmatory studies, Chabas and colleagues performed a microarray-based screen for genes differentially expressed in MS brain lesions and normal brain samples. Osteopontin and αB crystallin (Cryab), proteins abundantly expressed in MS lesions but not in normal brain, were singled out for investigation in a string of elegant mechanistic experiments that established (i) the role of these proteins in the pathogenesis of MS and (ii) their potential as therapeutic targets for the treatment of MS. These findings are now being translated into potential clinical applications: (1) monoclonal antibodies against osteopontin are being developed and evaluated for therapeutic efficacy in animal models of MS; (2) based on Cryab’s function as an antigen that elicits potent, protective B- and T-cell responses in MS patients, a novel therapeutic approach using recombinant Cryab for the treatment of relapsing remitting MS is being pursued.

Using high-throughput proteomic analysis of MS lesions, the same researchers identified protein signatures that distinguish between acute, chronic active, and chronic MS plaques. They found that a number of the proteins unique to chronic active plaques belong to the coagulation family of proteins, and they investigated the role of two of these proteins – tissue factor and protein C inhibitor (PCI) – in experimental autoimmune encephalitis (EAE), a mouse model of MS. Tissue factor promotes the formation of the proinflammatory mediator thrombin, and administration of a thrombin inhibitor was shown to ameliorate established EAE. Similarly, PCI inhibits the anticoagulant-activated protein C (aPC), and administration of recombinant aPC attenuated established EAE. Thus, reversal of events downstream of the proteins identified in MS plaques ameliorated autoimmune demyelination, implicating aberrant coagulation events in the pathogenesis of EAE/MS.

In the elegant approach illustrated in the study described above, hypothesis-generating genomic and proteomic experiments are performed to guide further experimental investigation. While the initial discovery step in the identification of a candidate drug target takes place in the pathologically transformed human tissue, in this case the MS plaque, accelerated “learning” about the target and its disease-specific mode of action is then staged in the appropriate animal model of disease (in this case EAE), and, coming full circle, the insights gleaned are swiftly applied to the treatment of human disease. Since the discovery steps take place in the human target tissue, the risk of identifying and selecting irrelevant molecular targets is greatly reduced.

With regard to clinical disease classification and monitoring, important developments include methods that increase the accuracy and reproducibility of measuring MS autoantibodies identified by genomic and proteomic approaches, and improvements in the alignment of genomic and proteomic findings, which we anticipate will contribute to the development of clinical assays for monitoring MS patients, as detailed below.

**Patient profiling to select therapy, monitor disease, and predict outcomes: development of multiplex assays**

The hypothesis currently being tested in multiple academic and industry laboratories worldwide is that molecular
signatures identified by genotyping, transcriptional profiling, or proteomic profiling experiments will provide greater utility for predicting clinically important endpoints and outcomes than do conventional methods. The success of a personalized medicine approach to the treatment of autoimmune disease will hinge on two main prerequisites: (1) robust (molecular) classification of clinically heterogeneous syndromes and (2) ability to stratify patient populations into relatively homogeneous subgroups with respect to clinically relevant outcomes.

**Single parameter versus multi-parameter assays**

A personalized medicine approach may be achieved through the measurement of a single genotypic/phenotypic variable, a few variables, or combinations of many variables that have predictive value. Variables may be molecular, imaging, or clinical parameters. Historically, only a few single biomarkers have proven useful for the prediction of disease severity at early stages; examples include anti-Scl-70 autoantibodies predictive of pulmonary disease in systemic sclerosis, maternal anti-Ro antibodies predictive of congenital heart block in newborns, and anti-double stranded DNA antibodies predictive of flares of lupus nephritis (reviewed by van Mühlen and Tan). Combining several markers has already been shown to improve the ability to predict outcomes in autoimmune diseases. For instance, combinations of autoantibody specificities are better able to predict the risk of developing autoimmune diabetes than is any single autoantibody specificity. Clinical immunologists believe that combinations of markers are needed for better prediction of long-term outcome in chronic autoimmune diseases such as RA and MS. Likewise, the multiple biomarker approach is of added value in predicting long-term outcome in chronic cardiovascular disease. Unsurprisingly, the market for multiplex assays is expected to be huge.

**Challenges in the introduction of multiplex assays into clinical practice**

Although replacement of existing assays with new, improved assays is appealing, experience shows that revisions of established diagnostic paradigms may take many years if not decades. A recent example in autoimmune disease is anti-citrulline antibody testing in RA, which emerged in the mid-1990s and subsequently competed with “classical” RF testing as the standard test for diagnosis of RA, and has only recently begun to displace RF testing in clinical decision-making. This gradual change in clinical practice was supported by a large number of studies in which an anticyclic citrullinated peptide (CCP) assay was used for the specific detection and quantification of anti-citrulline autoantibodies in thousands of patients, as well as metanalyses suggesting that CCP ELISAs are more useful for diagnostic and predictive purposes. For multiplex assays, the above development challenges are themselves “multiplexed”, as each individual parameter must be fully validated in a large number of patients before such assays can be approved for use in clinical decision making. To our knowledge, this hurdle has not yet been cleared by any multiplex test and no such assay has yet achieved health-authority approval for clinical use. Nevertheless, many promising assays are in exploratory development both in academia and in industry, and a few examples are provided below.

**Autoantibody profiling by antigen microarray**

Autoantibodies have proven to be the single most important category of biomarkers used in clinical decision making in autoimmunity and are typically measured by highly validated, gold-standard assays. While single-parameter autoantibody assays are established and valuable tools for the classification of many autoimmune diseases, a number of important clinical questions cannot be answered by these conventional assays alone. Testing panels of autoantibodies may greatly increase the clinical utility of autoantibody testing for a variety of clinical applications.

To address this unmet medical need, our lab has developed antigen microarrays for multiplexed, high-throughput profiling of biological samples from patients with autoimmune disease. Recently, these early 214-feature antigen arrays were substantially expanded and diversified into a spectrum of larger-scale antigen arrays to serve a variety of purposes. Types of arrays now include synovial antigen arrays, connective tissue disease and lupus antigen arrays, and myelin sheath antigen arrays. Antigen arrays contain protein and peptide antigens, lipid antigens, carbohydrate antigens, or mixtures thereof. Candidate antigens (eg, those identified in screens of RA synovial proteome by 2DE-DIGE mass spectrometry) and series of overlapping peptides derived from protein antigens are continually being added to the arrays. Moreover, critically important posttranslational modifications, such as citrullination or phosphorylation, can be analyzed by antigen microarrays, and citrullinated peptides are included on the arrays. Over the past seven or so years, we have used these antigen microarrays for (1) identification of novel antigen targets for specific immunotherapy, (2) identification of...
molecular profiles that classify heterogeneous autoimmune diseases, and (3) fine-epitope mapping of immune responses and epitope spreading in animal models of autoimmunity. Key discoveries include previously unknown epitopes targeted in RA and MS, autoantibody reactivity patterns associated with clinically important surrogate markers of disease progression, and combined autoantibody-cytokine signatures that allow the stratification of patients into clinically relevant subgroups. Current efforts are directed toward the use of antigen microarrays in conjunction with other high-throughput methods, such as bead-based assays, for the measurement of blood cytokines and other soluble markers in order to identify molecular signatures associated with response to anti-tumor necrosis factor (TNF) therapy. Experiments using samples from patients treated with other biological agents, eg, rituximab or abatacept/CTLA4-Ig, are being designed. Related efforts pertain to the use of arthritis antigen arrays to analyze longitudinally collected serum samples from predisease timepoints (described elsewhere in this paper).

High-throughput screening to generate comprehensive characterization of the autoantibody repertoire in SLE is now being pursued by several academic groups. In one such study, confirmed and expanded earlier results of differences in immunoglobulin G (IgG) and IgM autoantibody repertoires among patients with lupus, and these differences may be relevant in the clinical setting.

An interesting paper recently reported an association of baseline autoantibody profiles and elevated BLyS, a master regulator of many B-cell abnormalities in SLE patients, with shorter duration of disease remission and increased likelihood of disease exacerbation in patients with lupus nephritis undergoing therapy with rituximab-based B-cell depletion. These findings lend support to the hypothesis that, despite complete B-cell depletion, long-lived plasma cells committed to certain autoantibody specificities continue to produce autoantibodies, giving rise to immune complexes that drive the inflammatory process. Given the huge number of autoantibodies produced in patients with SLE, it is reasonable to speculate that large-scale antibody profiling in SLE may identify a pretreatment autoantibody signature with greater sensitivity for predicting inadequate response to rituximab, ultimately leading to robust identification of those rituximab-treated patients in need of additional immunosuppressive therapy.

Another important area of proteomic research in RA is the detection of molecular markers that precede the symptoms of full-blown clinical disease. While specific serum autoantibodies predate the clinical onset of type 1 diabetes (T1D) or SLE by years, little is known about molecular predictors of RA development in healthy individuals. Standard, low-throughput approaches to identify molecular markers associated with RA-related autoimmunity in healthy individuals at increased risk of developing RA may be ineffective due to the selection of a candidate marker molecule that is inadequate or not specific to the particular disease. Pilot studies suggested that anti-CCP autoantibodies and the IgM or IgA RF isotypes may be specific markers in healthy subjects at risk of developing RA years later, warranting further research. Elucidation of preclinical autoimmune states is extremely challenging as it requires valuable sera or tissue whose availability is limited, and reliable high-throughput assays. Such samples are available only through specialized blood or tissue banks, for example those containing samples from the Studies of the Etiology of RA (SERA) cohort, comprising 605 subjects who are parents of children at increased genetic risk of developing T1D. This cohort is particularly valuable since it is enriched for human leukocyte antigen (HLA)-DR4, a susceptibility marker for both T1D and RA. A second cohort of 622 first-degree relatives of patients with RA was recruited through rheumatology clinics and community outreach efforts at the University of Colorado Denver, Cedars–Sinai Medical Center, The Feinstein Institute for Medical Research, the Rheumatoid Arthritis Investigational Network (RAIN) from the University of Nebraska Medical Center, and the Benaroya Research Institute at Virginia Mason Arthritis Center, Seattle, WA. High-throughput autoantibody and cytokine profiling of these longitudinally collected serum samples would, under ideal conditions, circumvent the problem of small sample-volume while affording the opportunity to screen simultaneously for many molecules of interest. Studies using arthritis antigen microarrays and optimized bead-based cytokine assays, as described previously, were recently initiated using a US military cohort and a Dutch cohort of blood donors to explore the emergence of autoantibodies and cytokines in the blood of healthy individuals who went on to develop RA years later (W H Robinson, Kevin Deane, V Michael Holers, Cor Verweij; unpublished data). This approach will also provide an opportunity to examine spatiotemporal relationships between cytokine production and autoantibody production in the preclinical stages of early RA.

2DE-DIGE and mass spectrometry

The vast majority of clinical proteomic studies are performed using two-dimensional gel electrophoresis (2DE) as the
established method to delineate the proteome of biological fluids including plasma, serum, urine and cerebrospinal fluid. Many investigators use unfractionated plasma since blood is easy to collect, and the method requires very little sample. However, this approach has significant limitations for the identification of less abundant proteins, and “hits” are frequently confined to the well known, usual suspects such as the acute phase proteins fibrinogen, C-reactive protein, and haptoglobin. Although these highly abundant plasma proteins are markers with some clinical utility, changes in plasma levels of acute phase proteins likely lack specificity for the prediction of disease-specific clinical outcomes. Thus, robust measurements of less abundant plasma proteins may ultimately be more important for clinical prediction purposes.

Methods to identify less abundant proteins that are aberrantly expressed in the plasma of patients with autoimmune conditions include enrichment techniques that improve the lower limit of detection in classic 2DE-DIGE experiments. Enrichment is usually achieved by depletion of abundant proteins such as albumin and immunoglobulins by using affinity columns. In a study using serum samples obtained from 10 RA patients before and after anti-TNF therapy, Dwivedi and colleagues\(^48\) depleted serum of the 12 most abundant serum proteins and then performed mass spectrometry analysis of the serum proteome to identify a large number of TNF-regulated proteins. Samples enriched for less abundant proteins hold promise for the identification of tissue leakage products and other proteins whose plasma concentrations are typically in the low nanoliter and subnanoliter range.\(^49\) Major disadvantages of enrichment methods are the considerably larger sample volumes needed, the higher intersample variability of protein concentrations in the enriched samples, and the high costs, which together make this approach less tractable for clinical proteomics applications. Thus, efforts are underway to improve the quantitative analysis of the proteome in unfractionated plasma samples.

Analysis of unfractionated, paired samples collected longitudinally from as few as 20 subjects may yield statistically significant differences between groups in many more 2DE-DIGE spots than previously thought.\(^50\) The authors claim that both technical reproducibility and their proposed analysis algorithm allows differences in many more proteins to be accurately analyzed between groups, thus lending greater depth and breadth to gel-based analysis of the plasma proteome while preserving the advantages of using unfractionated plasma. Once available, reference plasma protein maps will further facilitate the interpretation and extrapolation of data from 2DE-DIGE experiments in autoimmunity, in a similar manner to those gel-based databases proposed almost 20 years ago by Anderson and colleagues\(^51\) in the context of drug effects at the proteomic level in drug toxicology studies. Of note, strict standardization and reporting procedures for plasma proteomics, such as those proposed by the HUPO consortium,\(^52\) will first need to be fully implemented before meaningful comparisons between autoimmune disease 2DE-DIGE datasets from different sources can be made.

The presence – in citrullinated form – of the candidate antigens fibrinogen, vimentin, and fibronectin in RA synovial fluid was recently confirmed by SDS-PAGE mass spectrometry analysis.\(^53\) Similarly, synovial alpha-enolase was identified by mass spectrometry and confirmed as a specific target of anti-citrulline autoantibody responses in RA.\(^54\) While these candidate antigens were described previously, newly identified, posttranslationally modified proteins may be added to the list of important proteins present in RA, but not control, synovium. Comprehensive mapping of citrullinated synovial proteins, and identification of proteins that trigger a specific and prominent immune response, may enable a tolerizing immunotherapy approach to the treatment of RA, the premise being that these modified proteins initiate and perpetuate the evolution of a maturing autoimmune response.

An important addition to the toolbox of -omics researchers is laser capture microdissection (LCM), a technique for isolating pure cell populations from a heterogeneous tissue section or cytological preparation through direct visualization of the cells in a high-throughput manner.\(^55\) This robotically controlled technique can be applied to molecular profiling of autoimmune diseases and should enable the correlation of a cell molecular profile with a cell population of interest. The use of LCM is particularly appealing in autoimmune diseases that involve multiple immune cell populations. The microdissected tissue may be used for DNA, RNA, or protein analyses.

**Genomics of peripheral blood cells**

Chaussabel and colleagues\(^56\) recently proposed an interesting modular approach to the discovery of clinically relevant biomarker signatures in patients with SLE. In this approach, genome-wide transcriptional profiles generated by microarray analysis of patients’ blood are organized into modules that are defined by genes coordinately expressed in multiple disease data sets. Changes in gene expression at the module level enabled the authors to design disease-specific transcriptional fingerprints. They propose the application of these modules to delineate a transcriptional
indicator of disease progression in patients with SLE. Although this approach is highly appealing and practicable, a well-known limitation of transcriptional profiling in SLE is the inability to capture the important posttranslational protein modifications that have been associated with both T-cell and B-cell dysfunction in SLE.\textsuperscript{57,58} Comprehensive characterization of the huge array of lupus neoantigens is therefore advocated for a comprehensive genomic and proteomic picture of SLE immunity (see also 2DE-DIGE and mass spectrometry above).

Improvements in labeling and detection techniques will allow more reliable microarray-based quantification of changes in serum autoantibody titers, as recently demonstrated for a number of known and newly discovered SLE autoantigens,\textsuperscript{59} and may facilitate the translation of antigen microarray technology to the clinic. More recently, label-free surface plasmon resonance biosensor systems that allow determination of binding kinetics and affinity constants arrays are being explored in this context.\textsuperscript{60}

Examining the transcriptome of a specific cell population such as blood leukocytes may inform the development of tests that enable more reliable and earlier diagnosis, as well as early assessment of response to immunosuppressive therapy. In a recent translational study, peripheral blood leukocytes from children with systemic onset juvenile idiopathic arthritis (SoJIA) were profiled by microarrays, and a specific 12-gene biomarker signature was identified that allowed early distinction between SoJIA and other febrile illnesses.\textsuperscript{61} Such leukocyte signatures, once validated in independent cohorts, have great potential for conversion into a diagnostic test to guide early immunosuppressive intervention in SoJIA, while other leukocyte signatures may be suitable to monitor response to anti-interleukin-1 therapy in SoJIA\textsuperscript{61} and potentially in SLE.\textsuperscript{62}

A different approach was taken by Nakou and colleagues,\textsuperscript{63} who performed transcriptional profiling of bone marrow cells derived from SLE patients. The authors compared gene expression profiles in bone marrow cells with those in peripheral blood mononuclear cells (PBMCs) and found that only gene signatures in bone marrow cells were able to cluster patients into two major subgroups, active and inactive disease. Although the number of patients investigated in this study was small, the findings indicate that, in complex autoimmune diseases such as SLE, PBMCs may not be the most useful surrogate cells in which to discover clinical biomarkers. This is in contrast to earlier findings, based on relatively larger patient populations, indicating the existence of disease-specific PBMC signatures, as well as “sentinel” PBMC signatures that are common to several autoimmune diseases.\textsuperscript{64}

**Expansion to new autoimmune disease areas**

Autoimmune diseases not previously investigated on a genomic or proteomic scale are now the subject of high-throughput screening studies and include scleroderma,\textsuperscript{65} systemic sclerosis,\textsuperscript{16} dermatomyositis,\textsuperscript{66} juvenile idiopathic arthritis,\textsuperscript{67} and Behçet’s disease.\textsuperscript{68} Studies on other less common diseases have been conducted, results of which have provided the first piece of evidence of distinct proteomic patterns in autoimmune diseases that otherwise exhibit quite similar phenotypic presentation.\textsuperscript{69} It is hoped that other less common but equally malignant autoimmune inflammatory conditions, including orphan diseases such as retroperitoneal fibrosis and relapsing polychondritis, will soon be subjected to genomic and proteomic investigations. New molecular insights gleaned from genomic and proteomic profiling of rare diseases will provide pharmaceutical companies with new incentives to test their pathway-specific therapies in this neglected field, ultimately leading to more and better therapeutic choices for patients with these serious conditions.

**Pharmacogenomics in autoimmunity**

A few interesting developments in the field of pharmacogenomics deserve mention. Studies are emerging that examine the proteome or transcriptome of target tissues or blood for markers of response to disease-modifying drugs (DMARD), eg, the first-of-a-kind \textit{ex vivo} study by Andreas and colleagues\textsuperscript{70} on changes of the RA chondrocyte transcriptome after DMARD therapy; a small serum proteome study demonstrating that a good clinical response to infliximab is associated with a 20% decrease in levels of each of a panel of 39 TNF-regulated serum proteins;\textsuperscript{46} and a study showing changes in gene expression in skin of chronic psoriasis patients undergoing immunosuppressive therapy.\textsuperscript{71} In this last study, the authors report that a two-pathway genetic signature – comprising the TH1 and TH17 pathways – in skin biopsies is associated with disease regression. Interestingly, the gene expression changes in response to cyclosporine A at a relatively early time point occurred in skin rather than blood, prompting the authors to speculate that these data may help to explain therapeutic activities in tissues that are not accessible to biopsy analysis. In another interesting study, transcriptional profiling was performed on peripheral blood of
16 RRMS patients with relapsing-remitting multiple sclerosis at baseline and one month after the start of βIFN therapy. A baseline signature of 15 βIFN regulated genes was identified that negatively correlated with clinical response at one, three, and six months of therapy with βIFN. Of note, the authors have validated and confirmed this candidate biomarker in an independent group of 30 RRMS patients.

Although systems biology studies are beyond the scope of this review, it should be noted that the reliability of databases used to build functional networks is continually improving, and thus systems biology studies are increasingly making their mark on the literature.  

**Genome-wide association studies (GWAS)**  
Before 2006, only a handful of non-HLA genetic disease associations were identified using the classical candidate gene approach and linkage analysis, tracing transmission of disease within families, or comparing frequencies of genetic variants between affected and unaffected individuals in larger populations (reviewed by Altshuler and colleagues). While successful to some extent, these studies proved inadequate to unravel complex genetic traits contributing to susceptibility in polygenetic disorders including autoimmune diseases. In the mid-nineties, a genome-wide approach to association studies was proposed, and about ten years later the first GWAS were published including studies of several autoimmune diseases.

The inflammatory bowel disease (IBD) field has since seen an explosion of new molecular data that are only beginning to be translated to clinical use. Most of this novel data come from multiple GWAS on Crohn’s disease (CD) that have significantly advanced our knowledge of the genetic landscape of IBD, outpacing progress in the identification of new risk alleles in other immune diseases. Thus, in this section, we will use CD as an example to outline the benefits and limitations of GWAS. Of note, the large number of risk alleles identified for CD so far is attributable to the fact that the rate of discoveries is correlated with both the magnitude of heritability and the number of patients scanned, with CD being among the autoimmune diseases with the highest heritability (sibling relative risk ratio [λs] = 30) and largest patient populations screened. While the landmark GWAS of 14,000 patients (including 2000 patients with CD) and 3000 control subjects, undertaken by the Wellcome Trust Case Control Consortium (WTCCC) of 50 British groups and published in 2007, introduced the wider medical community to the concept of “risk genotyping”, the very first GWAS of patients with CD identified IL23R as major susceptibility gene in IBD. Thus, of particular interest are the most recent replication studies that confirm the major risk alleles related to the IL12/23 pathway in CD, reported for an American cohort and a Dutch/Belgium cohort. However, in a noteworthy critique of the widespread over-hyping of the clinical utility of replicated SNPs with highly significant odds ratios for personalized medicine purposes, Jakobsdottir and colleagues emphasize that strong association (low P-value) by no means translates to accurate classification of cases and controls. In a model of five replicated risk alleles for CD, an AUC (area under the curve; a measure of discrimination of cases and controls) of only 0.66 was calculated, suggesting that these five SNPs are of only moderate value for the diagnosis of CD. Moreover, the 32 known risk alleles account for only one fifth of the heritability of CD, illustrating the challenge of comprehensively mapping genetic risk in this disease population. GWAS have contributed tremendously to our knowledge of the genetic contributors to disease etiology, and are thus important hypothesis-generating tools; in contrast, their added practical value for clinical prediction of risk in subjects with autoimmune disease remains small.

Important hypotheses await testing in clinical trials of patients with IBD who possess polymorphisms known to confer risk for the development of CD. For example, it could be speculated that a subgroup of patients with polymorphisms in the IL23R gene will have a better clinical response to biological therapies targeting this pathway than will patients who do not possess this polymorphism. Yet another hypothesis to be tested in prospective cohorts is whether SNP markers are associated with severity of disease or unfavorable long-term outcomes. If so, patients with baseline genotype profiles suggestive of risk for the later development of disease complications may qualify for more aggressive therapy early on.

Finally, research on CD provides an excellent example of GWAS corroborating earlier findings relating to mechanisms of disease. First, the innate immune system was implicated in the development of CD on the basis of the presence in certain CD patients of dysfunctional mutants of NOD2, a protein that interacts with intestinal microbes and is critical for the maintenance of mucosal homoeostasis and effective host defense. GWAS subsequently identified polymorphisms in the autophagy genes ATG16L1 and IRGM as susceptibility traits for CD; ATG16L1 and IRGM encode proteins involved in host defense against invasive intestinal pathogens, further supporting the hypothesis that the innate immune system plays a key role in CD.
Proteomic studies on IBD

While many of the less common genetic variants that confer additional genetic risk for the development of IBD still await identification, IL-12/23 pathway components that have been suggested as putative susceptibility genes will need to be scrutinized and confirmed at the protein level. The same holds true for other newly identified susceptibility genes, especially those common to several autoimmune diseases and thus potentially representing master regulators of autoimmunity (ie, PTNP22, IL-21, STAT4, IL-2RA, etc), as reviewed by Gregersen and Olsson. An excellent review of genomic and proteomic toolkits for clinical applications in IBD was recently published, and this topic will not be discussed in detail here. Although few and far between, proteomic studies addressing questions of pathway dysregulation in patients with IBD, or examining protein associations with clinical phenotypes, include serum proteome profiling for biomarker discovery in IBD, proteomic studies on prediction of response to infliximab in IBD, and a pilot study that interrogated proteome changes by 2DE-DIGE and MALDI-TOF mass spectrometry analysis of intestinal epithelial cells from CD and ulcerative colitis (UC) patients and compared proteomic profiles in inflamed versus non-inflamed epithelium from samples of the same patients. One of the interesting candidate proteins that was upregulated in inflamed epithelial cells from both CD and UC patients was Rho-GDP dissociation inhibitor R, while programmed cell death protein-8 was upregulated in UC patients.

Concluding remarks

Realization of the challenges that face the translation of findings from -omics experiments into clinical practice has led to the early excitement over clinical implications of GWAS being replaced by a more critical appraisal of the value of SNP associations for the prediction of disease risk. Proteomic analysis of biological fluids by 2DE-DIGE mass spectrometry is in its infancy and many commentators consider the roads leading to approval of biomarker assays long and winding. Nonetheless, exploratory genomic/proteomic studies have already proven their remarkable ability to shed light on the mechanisms of autoimmune disease and identify new therapeutic targets, as exemplified by the delineation of autophagy pathways in CD and by the highly innovative lipidomic and proteomic approaches used for the study of MS. Furthermore, microarray-based investigations have led to the discovery of several candidate markers that are now being validated and revalidated on industry-standard platforms, and to the development of tolerizing antigen therapeutics anticipated to enter clinical testing.

In summary, these are exciting times for researchers and clinicians ready to embrace the vision of personalized medicine, and it is hoped that translating promising findings from genomic and proteomic studies into both innovative therapies and superior clinical predictors will ultimately benefit those for whom it matters most, the patients with debilitating autoimmune diseases.

Disclosure

We are grateful to Dr L Steinman, VM Holers, and PJ Utz whose ideas, visions, and concepts influenced this review. We also appreciate the input of Dr TM Lindstrom in the editing of this manuscript. We apologize to those scientists in the field whose work has not been cited as a result of space limitations. W.H.R. was funded by NIH NHLBI contract N01 HV 28183, NIAMS R21-A1069160, NIAMS R01-AR-054822, and Veterans Affairs Health Care System.

References


76. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.


